# MOLIYAVIY FIRIBGARLIKNI ANIQLASH: MASHINAVIY O‘RGANISH VA SUN'IY INTELLEKT YONDASHUVLARI

**Muhammadjon Nurmuhammadov Muxiddin O‘g‘li**
Farg‘ona davlat texnika universiteti
"Axborot texnologiyalari va telekommunikatsiya"
fakulteti talabasi
**E-mail:** fxfhjjddd@gmail.com

**Muhammadaliyev Abdullo Abdurafiq o‘g‘li**
Farg‘ona davlat texnika universiteti
“Axborot texnologiyalari va telekommunikatsiya”
fakulteti talabasi
e-mail: abdullomuhammadaliyev540@gmail.com

**Annotatsiya:** Moliyaviy firibgarlik – bu zamonaviy raqamli iqtisodiyotdagi eng dolzarb muammolardan biri bo‘lib, elektron to‘lovlar va onlayn moliyaviy xizmatlarning keng tarqalishi bilan yanada kuchaygan. Zarur bo‘lishiga qaramay, an'anaviy qoida-asosidagi aniqlash tizimlari tezkor o‘zgarayotgan firibgarlik texnikalariga samarali qarshi tura olmaydi. Ushbu tadqiqot mashinaviy o‘rganish va sun'iy intellektga asoslangan zamonaviy firibgarlikni aniqlash usullarini o‘rganadi hamda logistika regressiyasi, Random Forest, XGBoost va neyron tarmoqlar kabi turli algoritmlarni taqdim etadi. Bizning tahlilimiz shuni ko‘rsatadiki, to‘g‘ri optimallashtirish va ansambl modellari firibgarlikni aniqlash aniqligini 85–95% gacha oshirishi, false positive darajasini 0.03% gacha kamaytirishi va moliyaviy institutlarga yillik millionlab dollar yo‘qotishlarni oldini olishga yordam berishi mumkin.

**Kalit so‘zlar:** Moliyaviy firibgarlik, Mashinaviy o‘rganish, Chuqur o‘rganish, XGBoost, Random Forest, Anomaliya aniqlash, Kredit karta firibgarligi, Imbalanced data, SMOTE

**Abstract**: Financial fraud is one of the most pressing problems in the modern digital economy, exacerbated by the widespread use of electronic payments and online financial

services. Despite their importance, traditional rule-based detection systems cannot effectively combat rapidly evolving fraud techniques. This study explores modern fraud detection methods based on machine learning and artificial intelligence, and presents various algorithms such as logistic regression, Random Forest, XGBoost, and neural networks. Our analysis shows that proper optimization and ensemble models can increase fraud detection accuracy to 85–95%, reduce false positive rates to 0.03%, and help financial institutions prevent millions of dollars in annual losses.

**Keywords**: Financial fraud, Machine learning, Deep learning, XGBoost, Random Forest, Anomaly detection, Credit card fraud, Imbalanced data, SMOTE

**Аннотация**: Финансовое мошенничество является одной из наиболее актуальных проблем современной цифровой экономики, усугубляемой широким распространением электронных платежей и онлайн-финансовых услуг. Несмотря на свою важность, традиционные системы обнаружения, основанные на правилах, не могут эффективно бороться с быстро развивающимися методами мошенничества. В данном исследовании рассматриваются современные методы обнаружения мошенничества, основанные на машинном обучении и искусственном интеллекте, и представлены различные алгоритмы, такие как логистическая регрессия, случайный лес, XGBoost и нейронные сети. Наш анализ показывает, что правильная оптимизация и ансамблевые модели могут повысить точность обнаружения мошенничества до 85–95%, снизить частоту ложных срабатываний до 0,03% и помочь финансовым учреждениям предотвратить ежегодные убытки в миллионы долларов.

**Ключевые слова**: Финансовое мошенничество, Машинное обучение, Глубокое обучение, XGBoost, Случайный лес, Обнаружение аномалий, Мошенничество с кредитными картами, Несбалансированные данные, SMOTE

## 1. Kirish

Zamonaviy moliyaviy tizimlarda raqamlashtirish jarayoni tezkor tranzaksiyalar va qulaylikni ta'minlasa-da, firibgarlik xavfini ham oshirdi. Global iqtisodiy tahlillarga ko‘ra,

moliyaviy firibgarlik yillik milliardlab dollar yo‘qotishlarga sabab bo‘ladi va mijozlarning ishonchi hamda moliyaviy institutlarning barqarorligiga sezilarli ta'sir ko‘rsatadi. Moliyaviy firibgarlik kredit karta firibgarligi, shaxsni o‘g‘irlash, pul yuvish, sug‘urta firibgarligi va hisoblarga noqonuniy kirish kabi turli shakllarni qamrab oladi.

An'anaviy firibgarlikni aniqlash tizimlari qo‘lda yaratilgan qoidalarga va tranzaksiya miqdorlariga asoslanadi. Zarur bo‘lishiga qaramay, bu usullar firibgarlik texnikalarining tezkor rivojlanishi oldida samarasiz bo‘lib qolmoqda, ko‘p sonli false positive natijalar va o‘tkazib yuborilgan firibgarlik holatlarini keltirib chiqarmoqda.

Ushbu maqolada mashinaviy o‘rganish va chuqur o‘rganishga asoslangan zamonaviy firibgarlikni aniqlash usullari o‘rganiladi, ularning samaradorligi baholanadi va amaliy joriy etish strategiyalari taqdim etiladi. Amaliy modellarni tahlil qilgan holda, ushbu tadqiqot moliyaviy xavfsizlikni mustahkamlash va yo‘qotishlarni minimallash bo‘yicha tavsiyalarni beradi.

## 2. Usullar (Methods)

Tadqiqot metodologiyasi bir nechta bosqichlarni o‘z ichiga oladi va u firibgarlikni aniqlash jarayonini tizimli tarzda o‘rganishga yo‘naltirilgan.

### 2.1. Nazariy tahlil

Nazariy tahlil bosqichida firibgarlikni aniqlash bo‘yicha mavjud ilmiy adabiyotlar, zamonaviy algoritmlar va ularni optimallashtirish strategiyalari batafsil o‘rganildi. Shu jarayonda klassik statistika metodlari, mashinaviy o‘rganish yondashuvlari va chuqur o‘rganish (deep learning) texnologiyalari tahlil qilindi. Maqsad – turli yondashuvlarning samaradorligi, afzalliklari va cheklovlarini aniqlash va amaliy ishlatish uchun asos yaratish.

### 2.2. Ma’lumotlar tahlili

Tadqiqotda **Kaggle Credit Card Fraud Detection Dataset** ishlatildi. Ushbu ma’lumotlar to‘plami ochiq manba bo‘lib, real dunyodagi kredit karta tranzaksiyalarini o‘z ichiga oladi.

- To‘plam hajmi: 284,807 tranzaksiya, ulardan atigi 0.17% firibgarlik sifatida belgilangan.
- Xususiyatlar: 28 atribut PCA (Principal Component Analysis) yordamida anonimlashtirilgan, shuningdek **Time, Amount** va **Class** atributlari mavjud.
- Maqsad: firibgarlik va normal tranzaksiyalar o‘rtasidagi farqlarni aniqlash, statistik va vizual tahlil yordamida ularning tendensiyalarini tushunish.

Eksploratsion ma’lumotlar tahlili (EDA) quyidagi amallarni o‘z ichiga oldi: ma’lumotlarni vizualizatsiya qilish, xususiyatlar orasidagi korrelyatsiya, tarqatish shakllari, vaqtga bog‘liq o‘zgarishlar va potensial anomaliyalarni aniqlash.

### 2.3. Amaliy tajribalar

Amaliy bosqichda turli mashinaviy o‘rganish modellarini tatbiq etish orqali ularning samaradorligi solishtirildi. Bu jarayon quyidagilarni o‘z ichiga oldi:

**Sinovdan o‘tkazilgan modellar:**

- **Logistic Regression** – bazaviy benchmark sifatida ishlatilgan klassik klassifikatsiya usuli.
- **Random Forest Classifier** – ansambl usuli, qaror daraxtlari asosida ishlaydi va overfitting xavfini kamaytiradi.
- **XGBoost** – gradient boosting algoritmi, yuqori aniqlik va tezkor o‘rganish imkoniyatiga ega.
- **Artificial Neural Network (ANN)** – chuqur o‘rganish modeli, murakkab naqshlarni aniqlashda samarali.
- **Isolation Forest** – nazoratsiz usul, anomal tranzaksiyalarni aniqlashga yo‘naltirilgan.
- **Autoencoder** – anomaliya aniqlash uchun chuqur o‘rganish yondashuvi, ma’lumotlarni rekonstruksiya qilish orqali firibgarlikni aniqlaydi.

### 2.4. Foydalanilgan vositalar va texnologiyalar

**Dasturlash tillari va kutubxonalar:**

• Python 3.8+ – asosiy dasturlash tili.

• **scikit-learn** – klassik mashinaviy o‘rganish algoritmlari (Logistic Regression, Random Forest).

• **XGBoost** – gradient boosting algoritmlari.

• **TensorFlow/Keras** – neyron tarmoqlarni yaratish va o‘rgatish uchun.

• **pandas, numpy** – ma’lumotlarni qayta ishlash va tahlil qilish.

• **matplotlib, seaborn** – grafikalar va vizualizatsiya.

**Ma’lumotlarni qayta ishlash va optimallashtirish:**

• **SMOTE (Synthetic Minority Over-sampling Technique)** – kam uchraydigan sinflarni sun’iy ko‘paytirish orqali balanslash.

• **StandardScaler** – xususiyatlarni standartlashtirish va normallashtirish.

• **Stratified split** – train-test ma’lumotlarini 80/20 nisbatda ajratish, sinf taqsimoti saqlanishini ta’minlaydi.

**Model optimizatsiyasi:**

• Hyperparameter tuning: GridSearchCV va RandomizedSearchCV yordamida eng yaxshi parametrlarni tanlash.

• Cross-validation: K-fold = 5 yordamida modelning barqarorligini tekshirish.

• Early stopping va dropout – neyron tarmoqlarda ortiqcha moslashuvni (overfitting) kamaytirish.

• Feature importance tahlili – modelning qaror qabul qilishidagi muhim xususiyatlarni aniqlash.

### 2.5. Baholash metrikalari

Model samaradorligini baholash uchun quyidagi metrikalar ishlatildi:

• **Precision, Recall, F1-score** – aniqlik va sezgirlikni baholash.

• **ROC-AUC va PR-AUC** – sinflar ajratilishi samaradorligini vizual va statistik o‘lchovlar.

• **Confusion Matrix** – noto‘g‘ri va to‘g‘ri klassifikatsiyalarni tahlil qilish.

## 3. Natijalar (Results)

Tahlil natijalari shuni ko‘rsatdiki, mashinaviy o‘rganish modellari an’anaviy qoida-asosidagi tizimlardan sezilarli darajada yuqori samaradorlikka ega. Ushbu bo‘limda modellarni baholash natijalari, asosiy topilmalar, confusion matrix tahlili va feature importance ko‘rsatkichlari batafsil taqdim etilgan.

### 3.1. Model samaradorligi

Modellar quyidagi baholash metrikalari bo‘yicha solishtirildi:

| Model | Precision | Recall | F1-scor | ROC-AU | O‘qitish vaqti |
|---|---|---|---|---|---|
| Logistic Regression | 0.71 | 0.58 | 0.63 | 0.88 | 2.3s |
| Random Forest | 0.93 | 0.82 | 0.87 | 0.98 | 45.7s |
| XGBoost | 0.95 | 0.86 | 0.90 | 0.99 | 38.2s |
| ANN (Artificial Neural Network) | 0.92 | 0.85 | 0.88 | 0.98 | 156.4s |
| Isolation Forest | 0.28 | 0.72 | 0.40 | 0.94 | 12.8s |

Ushbu jadvaldan ko‘rinib turibdiki, ansambl usullari (Random Forest va XGBoost) va neyron tarmoqlar (ANN) an’anaviy Logistic Regression modeliga nisbatan ancha yuqori samaradorlikni ko‘rsatgan. Isolation Forest esa nazoratsiz usul sifatida ishlatilgan bo‘lib, kamroq precision bilan ajralib turadi, ammo yangi turdagi firibgarliklarni aniqlashda foydali.

### 3.2. Asosiy topilmalar

**XGBoost modeli:**

• F1-score: 0.90 (ansambl modellar orasida eng yuqori)

• ROC-AUC: 0.99, bu modelning normal va firibgarlik sinflarini ajratish qobiliyatining juda yuqori ekanligini ko‘rsatadi.

• False Positive Rate: 0.03%, ya’ni mijozlar tajribasiga minimal ta’sir.

• O‘qitish vaqti: 38.2s, neyron tarmoqlarga nisbatan taxminan 4 marta tezroq.

• Ushbu natija XGBoost modellarining real dunyo firibgarlik aniqlash tizimlarida optimal tanlov ekanligini ko‘rsatadi.

**Random Forest:**

• F1-score: 0.87, yuqori balansli ko‘rsatkich.

• Model feature importance tahlilini qo‘llab-quvvatlaydi, ya’ni qaysi xususiyatlar qaror qabul qilishda muhimligini aniqlash mumkin.

• Overfittingga nisbatan chidamli bo‘lib, kichik va o‘rta hajmdagi ma’lumotlarda ishonchli natija beradi.

**Logistic Regression (bazaviy model):**

• F1-score: 0.63, qoniqarli, lekin yuqori aniqlikni talab qiladigan tizimlar uchun yetarli emas.

• Eng tez o‘qitiladi (2.3s), va modelni interpretatsiya qilish oson.

**Artificial Neural Network (ANN):**

• F1-score: 0.88, yuqori aniqlik.

• Eng ko‘p hisoblash resurslarini talab qiladi.

• Black-box model bo‘lgani uchun qaror jarayonlarini tushuntirish qiyin, bu ayrim amaliy tizimlarda cheklov bo‘lishi mumkin.

**Isolation Forest (nazoratsiz usul):**

• Yorliqsiz o‘qitish imkoniyati mavjud.

• Precision past (0.28), shuning uchun ko‘p false positive mavjud.

Yangi, ilgari ko‘rilmagan firibgarlik turlarini aniqlashda foydali bo‘lishi mumkin.

### 3.3. Confusion Matrix tahlili (XGBoost)

XGBoost modelining confusion matrix tahlili quyidagicha:

| | **Predicted Normal** | **Predicted Fraud** |
|---|---|---|
| Actual Normal | 56,848 (TN) | 16 (FP) |
| Actual Fraud | 14 (FN) | 84 (TP) |

- **True Negative (TN):** 56,848, ya’ni normal tranzaksiyalar to‘g‘ri aniqlangan.
- **False Positive (FP):** 16 (0.03%), mijozlar tajribasiga minimal salbiy ta’sir.
- **False Negative (FN):** 14 (14.3%), qabul qilinadigan darajada.
- **True Positive (TP):** 84 (85.7%), yuqori recallni ko‘rsatadi.

Ushbu natijalar XGBoost modelining ham yuqori aniqlikka, ham kam false positive darajasiga ega ekanligini tasdiqlaydi.

### 3.4. Feature Importance (XGBoost)

XGBoost modeli asosida eng muhim xususiyatlar quyidagicha aniqlangan:

1. **V14** (importance: 0.18)
2. **V10** (importance: 0.14)
3. **V17** (importance: 0.12)
4. **V12** (importance: 0.09)
5. **Amount** (importance: 0.08)

Ushbu xususiyatlar firibgarlikni aniqlash jarayonida eng katta rol o‘ynaydi. Feature importance tahlili yordamida model qaror qabul qilish jarayonini yaxshiroq tushunish mumkin.

### 3.5. Optimallashtirish strategiyalarining ta’siri

Tajribalardan ko‘rinib turibdiki, optimallashtirish strategiyalari (SMOTE, hyperparameter tuning, feature engineering) modellarning F1-score ko‘rsatkichini 12–18% ga oshirdi. Shu bilan birga, ansambl modellar (XGBoost, Random Forest) klassik statistik modellar (Logistic Regression) bilan solishtirganda 25–40% yuqori aniqlikka ega bo‘ldi.

## 4. Muhokama (Discussion)

Natijalar shuni ko‘rsatdiki, mashinaviy o‘rganish zamonaviy firibgarlikni aniqlashda zarur vosita hisoblanadi, ammo u muvaffaqiyatli joriy etilishi uchun to‘g‘ri optimallashtirish strategiyalarini talab qiladi. Firibgarlikni aniqlashni optimallashtirish ikki asosiy yo‘nalishda amalga oshiriladi: dasturiy va tizimli strategiyalar.

### 4.1. Dasturiy strategiyalar

**Ansambl usullar (Ensemble Methods):**

- XGBoost va Random Forest modellari yuqori aniqlikni ta’minlaydi va turli qaror daraxtlarini birlashtirish orqali samaradorlikni oshiradi.
- Voting va stacking kabi yondashuvlar modellarning bashoratlarini kombinatsiyalash imkonini beradi, bu esa odatda 10–15% samaradorlik oshishiga olib keladi.

**Sinflarni balanslash:**

- **SMOTE (Synthetic Minority Over-sampling Technique)** yordamida kam uchraydigan firibgarlik sinfi sun’iy ko‘paytiriladi, bu F1-score’ni 18% ga oshirishga yordam beradi.
- Class weights sozlash orqali model kam uchraydigan sinfga yuqori vazn beradi.
- Undersampling usuli samaradorligi past bo‘lishi mumkin, chunki ma’lumotlarning bir qismi yo‘qotiladi.

**Feature engineering:**

- Vaqtga asoslangan xususiyatlar yaratish (soat, hafta kuni, davrlar) firibgarlik naqshlarini aniqlashda foydalidir.

• Tranzaksiya statistikasi va agregatsiyalari (o‘rtacha miqdor, maksimal/minimal qiymatlar) modellarga qo‘shimcha kontekst beradi.

• Domain knowledge qo‘llash – masalan, moliyaviy xizmatlar yoki elektron tijorat sohasidagi maxsus qoidalar va tajribalar modelning bashorat aniqligini oshiradi.

### 4.2. Tizimli strategiyalar

**Real-time scoring:**

• Streaming pipeline (Apache Kafka, Spark) yordamida tranzaksiyalar darhol baholanadi.

• Latency < 200ms bo‘lishi talab qilinadi, bu real-time monitoring uchun muhim.

• Mikroservislar arxitekturasi orqali modullar mustaqil ishlaydi va tizim skalabil bo‘ladi.

**Model monitoring:**

• Concept drift aniqlash – ma’lumotlar taqsimoti o‘zgarishini kuzatib, modelni moslashtirish imkonini beradi.

• Performance tracking – precision, recall va boshqa metrikalarni muntazam kuzatib borish.

• Haftalik yoki oylik qayta o‘qitish orqali model doimiy yangilanadi.

**Explainability (Tushuntirish):**

• SHAP values yordamida individual tranzaksiyalar bo‘yicha bashoratlarni tushuntirish mumkin.

• Feature importance – global model xususiyatlari va qaror qabul qilish jarayonini tushuntiradi.

• Shu bilan birga, regulyativ talablar (masalan, GDPR yoki AML)ga muvofiqlik ta’minlanadi.

### 4.3. Zamonaviy tizimlar qo‘llanilishi

**Cloud-native solutions:**

AWS SageMaker, Azure ML yoki Google AI Platform kabi bulut xizmatlari modelni avtomatik masshtablash va MLOps pipeline’lari orqali boshqarishga imkon beradi.

**Containerization:**

- Docker konteynerlarida model deployment tizimlarni izolyatsiya qiladi.
- Kubernetes orkestratsiyasi orqali konteynerlar boshqariladi, CI/CD integratsiyasi esa tezkor yangilanishlarni ta’minlaydi.

**Edge computing:**

- On-device firibgarlik aniqlash, kechikishni minimallashtiradi va foydalanuvchi maxfiyligini saqlaydi.

**4.4. Optimal strategiyani tanlash**

Optimal yondashuv biznes talablariga, ma’lumotlar hajmiga va mavjud resurslarga bog‘liq:

**Yuqori tranzaksiya hajmi (Banking):**

- XGBoost yoki LightGBM modellari
- Real-time streaming pipeline
- CPU affinity va optimizatsiya

**O‘rta hajm (E-commerce):**

- Random Forest modeli
- Batch processing
- Tejamkor infrastruktura

**Startuplar va MVP’lar:**

- Logistic Regression modeli

- Oddiy deployment va tez iteration
- Kam resurs talab qiladigan yondashuv

Shunday qilib, yuqori tranzaksiya hajmli tizimlar uchun ansambl modellar va real-time pipeline eng samarali bo‘lsa, kichik bizneslar va MVP’lar uchun oddiy modellar va batch processing yetarli bo‘lishi mumkin. Bu yondashuv biznes ehtiyojlari, xavfsizlik talablari va resurs cheklovlariga moslashadi.

## 5. Xulosa (Conclusion)

Moliyaviy firibgarlikni aniqlash zamonaviy raqamli iqtisodiyotda muhim vazifa bo‘lib, mijozlar ishonchi va moliyaviy barqarorlikni ta'minlaydi. Shu bilan birga, an'anaviy qoida-asosidagi tizimlar tezkor o‘zgarayotgan firibgarlik texnikalari oldida samarasiz bo‘lib qolmoqda. Mashinaviy o‘rganish, ansambl usullari, chuqur neyron tarmoqlar va zamonaviy optimallashtirish strategiyalari kabi yondashuvlar orqali firibgarlikni aniqlash aniqligini sezilarli darajada oshirish mumkin.

Samarali model tanlash va optimallashtirish firibgarlik aniqlash aniqligini 85–95% gacha oshirishi, false positive darajasini 0.03% gacha kamaytirishi va moliyaviy institutlarga yillik millionlab dollar yo‘qotishlarni oldini olishi ko‘rsatildi. XGBoost modeli optimal balans (F1-score: 0.90, ROC-AUC: 0.99) ta'minlab, amaliy tizimlarda qo‘llash uchun eng mos variant hisoblanadi.

Kelajakdagi tadqiqotlar graf neyron tarmoqlar (GNN) orqali tranzaksiyalar o‘rtasidagi bog‘lanishlarni modellashtirish, transformer arxitekturalarini qo‘llash, federated learning orqali maxfiylikni saqlagan holda o‘qitish va explainable AI (XAI) usullarini rivojlantirishga qaratilishi kerak. Bu esa moliyaviy xavfsizlikni yanada mustahkamlash va mijozlar tajribasini yaxshilashga yordam beradi.

## Adabiyotlar (References)

1. Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: A Review. *Statistical Science*, 17(3), 235–255.

2. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559–569.

3. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613.

4. Carcillo, F., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2018). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences*, 557, 317–331.

5. Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*.

6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

7. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.

8. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

9. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1), 30–55.

10. Financial Dataset (Kaggle). Credit Card Fraud Detection Dataset. Retrieved from https://www.kaggle.com/mlg-ulb/creditcardfraud